



# Raman spectroscopy combined with a support vector machine for differentiating between feeding male and female infants mother's milk

RAHAT ULLAH,<sup>1,\*</sup> SARANJAM KHAN,<sup>1</sup> SAMINA JAVAID,<sup>2</sup> HINA ALI,<sup>1</sup>  
MUHAMMAD BILAL,<sup>1</sup> AND MUHAMMAD SALEEM<sup>1</sup>

<sup>1</sup>Agri-Biophotonics Division, National Institute of Lasers and Optronics (NILOP), Nilore, Islamabad 45650, Pakistan

<sup>2</sup>Department of Biotechnology, International Islamic University Islamabad (IIUI), Pakistan

\*rahat\_khalil@yahoo.com

**Abstract:** This study presents differentiation in milk samples of mother's feeding male and female infants using Raman spectroscopy combined with a support vector machine (SVM). Major differences have been observed in the Raman spectra of both types of milk based on their chemical compositions. Overall, it has been found that milk samples of mother's having a female infant are richer in fatty acids, phospholipids, and tryptophan. In contrast, milk samples of mother's having a male infant contain more carotenoids and saccharides. Principal component analysis and SVM further highlighted the differences between the two groups on the basis of differentiating features obtained from their Raman spectra. The SVM model with two different kernels, i.e. polynomial kernel function (order-2) and Gaussian radial basis function (RBF sigma-2), are used for gender based milk differentiations. The performance of the proposed model in terms of accuracy, precision, sensitivity, and specificity using the polynomial kernel function of order-2 have been found to be 86%, 88%, 85% and 88%, respectively.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

**OCIS codes:** (300.0300) Spectroscopy; (300.6450) Spectroscopy, Raman.

## References and links

1. O. Ballard and A. L. Morrow, "Human milk composition: Nutrients and bioactive factors," *Pediatr. Clin. North Am.* **60**(1), 49–74 (2013).
2. D. Tomé, J. Schwarz, N. Darcel, and G. Fromentin, "Protein, amino acids, vagus nerve signaling, and the brain," *Am. J. Clin. Nutr.* **90**(3), 838S–843S (2009).
3. K. Hinde and L. A. Milligan, "Primate milk: Proximate mechanisms and ultimate perspectives," *Evol. Anthropol.* **20**(1), 9–23 (2011).
4. S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomed. Opt. Express* **7**, 2249 (2016).
5. S. Marouzi, A. Sharifi Rad, S. Beigoli, P. Teimoori Baghaee, R. Assaran Darban, and J. Chamani, "Study on effect of lomefloxacin on human holo-transferrin in the presence of essential and nonessential amino acids: Spectroscopic and molecular modeling approaches," *Int. J. Biol. Macromol.* **97**, 688–699 (2017).
6. K. Duchén and L. Thorell, "Nucleotide and polyamine levels in colostrum and mature milk in relation to maternal atopy and atopic development in the children," *Acta Paediatr.* **88**(12), 1338–1343 (1999).
7. H. L. Månsson, R. Fondén, and H. E. Pettersson, "Composition of Swedish dairy milk," *Int. Dairy J.* **13**(6), 409–425 (2003).
8. P. Parodi, "Milk fats in human nutrition," *Aust. J. Dairy Technol.* **59**, 3–59 (2004).
9. S. M. Innis, "The Role of Dietary n-6 and n-3 Fatty Acids in the Developing Brain," *Dev. Neurosci.* **22**(5-6), 474–480 (2000).
10. W. Heine, M. Radke, and K. D. Wutzke, "The significance of tryptophan in human nutrition," *Amino Acids* **9**(3), 91–205 (1995).
11. K. Hinde, M. L. Power, and O. T. Oftedal, "Rhesus macaque milk: magnitude, sources, and consequences of individual variation over lactation," *Am. J. Phys. Anthropol.* **138**(2), 148–157 (2009).
12. R. K. Chen, L. W. Chang, Y. Y. Chung, M. H. Lee, and Y. C. Ling, "Quantification of cow milk adulteration in goat milk using high-performance liquid chromatography with electrospray ionisation mass spectrometry," *Rapid Commun. Mass Spectrom.* **18**, 1167–1171 (2004).
13. G. Maccabiani, E. Pavoni, M. Tilola, E. Agnelli, M. Simoni, F. D'Abrosca, and P. Boni, "Setting-up a PCR-based method for species identification in milk products," *Vet. Res. Commun.* **29**(S2 Suppl 2), 327–329 (2005).

14. S. Khorsand Ahmadi, M. Mahmoodian Moghadam, P. Mokaberi, M. Reza Saberi, and J. Chamani, "A comparison study of the interaction between  $\beta$ -lactoglobulin and retinol at two different conditions: spectroscopic and molecular modeling approaches," *J. Biomol. Struct. Dyn.* **33**(9), 1880–1898 (2015).
15. M. M. Moghaddam, M. Pirouzi, M. R. Saberi, and J. Chamani, "Comparison of the binding behavior of FCCP with HSA and HTF as determined by spectroscopic and molecular modeling techniques," *Luminescence* **29**(4), 314–331 (2014).
16. K. Saranjam, U. Rahat, M. Saleem, M. Bilal, R. Rashad, and K. Inamullah, "Raman spectroscopy analysis of dengue virus infection in human blood sera," *J. Opt.* **127**(4), 2086–2088 (2016).
17. S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomed. Opt. Express* **7**(6), 2249–2256 (2016).
18. C. Krafft, G. Steiner, C. Beleites, and R. Salzer, "Disease recognition by infrared and Raman spectroscopy," *J. Biophotonics* **2**(1-2), 13–28 (2009).
19. B. F. J. Manly, "Multivariate Statistical Methods: A Primer," 3rd Edition, (Chapman and Hall) 224–234, (2005).
20. R. Ullah, S. Khan, A. Khan, M. Saleem, H. Ali, M. Bilal, and M. Ahmed, "Infant gender-based differentiation in concentration of milk fats using near infrared Raman spectroscopy," *J. Raman Spectrosc.* **48**(3), 363–367 (2017).
21. M. F. Mrozek and M. J. Weaver, "Detection and identification of aqueous saccharides by using surface-enhanced Raman spectroscopy," *Anal. Chem.* **74**(16), 4069–4075 (2002).
22. C. Barbosa-Cesnik, K. Schwartz, and B. Foxman, "Lactation mastitis," *JAMA* **289**(13), 1609–1612 (2003).
23. M. A. Conlon and A. R. Bird, "The impact of diet and lifestyle on gut microbiota and human health," *Nutrients* **7**(1), 17–44 (2014).
24. B. Koletzko and M. Rodriguez-Palmero, "Polyunsaturated fatty acids in human milk and their role in early infant development," *J. Mammary Gland Biol. Neoplasia* **4**(3), 269–284 (1999).
25. S. M. Innis, "Dietary omega 3 fatty acids and the developing brain," *Brain Res.* **1237**, 35–43 (2008).
26. R. M. El-Abassy, P. J. Eravuchira, P. Donfack, B. Kammer, and A. Matey, "Fast determination of milk fat content using Raman spectroscopy," *J. Vibrational Spectroscopy* **56**(1), 3–8 (2011).
27. I. Stefanov, V. Baeten, B. De Baets, and V. Fievez, "Towards combinatorial spectroscopy: the case of minor milk fatty acids determination," *Talanta* **112**, 101–110 (2013).
28. A. Tankiewicz, D. Dziemiańczyk, P. Buczek, I. J. Szarmach, S. Z. Grabowska, and D. Pawlak, "Tryptophan and its metabolites in patients with oral squamous cell carcinoma: preliminary study," *Adv. Med. Sci.* **51**(Suppl 1), 221–224 (2006).
29. J. D. Carver, B. Pimentel, W. I. Cox, and L. A. Barnes, "Dietary nucleotide effects upon immune function in infants," *Pediatrics* **88**(2), 359–363 (1991).
30. C. L. Sánchez, J. Cubero, J. Sánchez, B. Chancón, M. Rivero, A. B. Rodríguez, and C. Barriga, "The possible role of human milk nucleotides as sleep inducers," *Nutr. Neurosci.* **12**(1), 2–8 (2009).
31. S. A. Atkinson, G. H. Anderson, and M. H. Bryan, "Human milk: comparison of the nitrogen composition in milk from mothers of premature and full-term infants," *Am. J. Clin. Nutr.* **33**(4), 811–815 (1980).
32. E. Schlimme, D. Martin, and H. Meisel, "Nucleosides and nucleotides: natural bioactive substances in milk and colostrum," *Br. J. Nutr.* **84**(S1 Suppl 1), S59–S68 (2000).

## 1. Introduction

Mother feed is recommended as a vital source of nutrition during infancy. Milk is a complex biological fluid which contains valuable components such as proteins, lipids, carbohydrates, vitamins and minerals which play major roles in infant's growth, immune system development and gastrointestinal maturation [1–3]. Moreover, it provides essential amino acids for the synthesis of hormones, enzymes, antibodies, glutathione, nucleotides (AGCT) and neurotransmitters [4,5]. Based on these bioactive components, variation in mother milk come up both within and between species depending upon factors like lactation stage, milk expression and maternal diet [6–8]. Similarly, fatty acids which are considered as a source of energy in milk vary in its composition depending upon the dietary intake and infant gender [9, 10]. Milk fat globules are composed of triacylglycerol (TAG) core surrounded by a complex membrane i.e. milk fat globule membrane (MFGM). Approximately 90% of the MFGMs are composed of polar lipids and proteins, whereas the remaining part includes glycoproteins, cholesterol, enzymes and other bioactive components [11].

Considering the significance of milk as a natural diet for infants, complete knowledge of the nutritional values of milk is a constant source of interest for both powdered milk manufacturers as well as consumers. For such purpose, various chemical analysis including gel-electrophoresis, mass spectrometry, enzyme linked immunosorbant assay (ELISA), polymerase chain reaction (PCR), nuclear magnetic resonance (NMR) and high performance liquid chromatography (HPLC) have been used for the analysis of mother milk [12,13]. The elemental as well as compositional analysis of milk protein, fats, vitamins and minerals

including magnesium, calcium, sodium, potassium, phosphate, nitrate and fatty acids etc. are mostly performed with the above mentioned techniques. These techniques require professional or skilled personnel's as well as are costly and time consuming.

In recent years, optical screening/diagnosis in combination with machine learning techniques have been widely used for characterizing biological media which gives reliable results [14,15]. Raman spectroscopic technique has the potential to distinguish among different biofluids based on their molecular composition as well as variation in concentrations of the biomolecules such as lipids and proteins [4, 16,17]. In the present study, applicability of Raman spectroscopy combined with support vector machine has been explored for differentiating human mother's milk feeding male and female infant. To the knowledge of author such study have not been published or reported elsewhere.

## 2. Materials and methods

### 2.1 Sample collection and preparation

Milk samples from 190 well-nourished feeding mothers from different areas of Rawalpindi, Pakistan have been used in current study. Among these, half samples were collected from lactating mothers feeding male infants with age group 1 to 15 months. Similar numbers of samples were also collected from lactating mothers feeding female infant with similar age group. All these samples were collected in the capped glass tube, after taking written consent of the feeding mothers. Ethical standards have been strictly followed for conducting this research study. Collection as well as experiment of all samples has been carried out after obtaining written approval from the ethical committee of PAEC General Hospital Islamabad. After collection, all milk samples were stored at low temperature ( $-16^{\circ}\text{C}$ ) till further use.

### 2.2 Raman spectrum acquisition

A 30  $\mu\text{l}$  droplet from each milk sample was put on a glass slide and ten Raman spectra were recorded from different position of each sample. Raman spectrometer ( $\mu\text{Ramboss DONGWOO OPRTON}$ , South Korea) with a spectral resolution of  $4\text{ cm}^{-1}$  was used in current study. A continuous wave (CW) excitation laser source of wavelength 532 nm was used for recording the Raman scattering from each sample. A microscope objective with numerical aperture of 0.7 and magnification of 100x was used both for focusing purpose as well as collection of Raman scattered light. An acquisition time of 10 seconds, with the laser power of 40 mW on sample surface was used for recording each spectrum. Raman spectral range was set from  $600\text{ cm}^{-1}$  to  $1800\text{ cm}^{-1}$  as this range contains useful information regarding protein, fatty acids etc [18]. The schematic diagram of the experimental set up has been explained previously [4].

### 2.3 Pre-processing and data analysis

In Raman spectra of all milk samples, fluorescence exists due to the presence of natural fluorophores. Apart from the fluorescence background, contribution from different noise sources also exist. So, before making spectral analysis, pre-processing is therefore necessary for the removal of different types of background signals that contribute to Raman spectra. Initially, all Raman spectra have been smoothened using Savitzky-Golay filters with five points taken as window through 3rd order polynomial fitting. All smoothened spectra are then background adjusted using in built 'msbackadj' function in Matlab.

### 2.4 Dimensionality reduction and support vector machine (SVM)

Principal component analysis (PCA) is generally used for identifying patterns in the data set, and expressing the data by highlighting their similarities and differences. In case of a high dimensional data, PCA is able to reduce the dimensionality by transforming the data into a new co-ordinate system called the principal component (PC) space, whereby each point has

new (x,y) values. PCA is considered as a dimensionality reduction technique, because most of the time, only the first few components explain more than 95% of the variance in the data. Consequently, one can disregard higher PC's without losing much of the information [19].

SVM is a data classification technique that acts in supervised way on the data set. It has many applications in biophotonics and pattern recognition [19]. SVM is mainly used in problems where the data are not linearly separable. For the classification of such a lineally inseperable data set, data are transformed to a higher dimensional space with the help of transforation functions (kernel functions), where the data might be linearly separable. The most commonly used kernels functions are polynomial and Gaussian radial basis function. Each of these kernel can be used with the different optimization methods such as sequential minimal optimization, quadratic programming and least square. In this way it not only classify the data but also optimizes the decision boundary by maximizing the margin between clusters of data [17].

For the statistical analysis of milk data, an SVM model has been developed. The model has been trained on the known milk data of both genders. Such algorithm basically reads an entire data set (whole Raman spectrum) and consider only the attractive features from the whole data sets. The same features is then used for determining unknown samples. Initially only first two principal compenets have been used, which hardly classify two types of milk clearly as shown in Fig. 4. In our previous work, PCA separated the milk Raman data sets of different infant genders very well because those samples were collected only from specific age group i.e. 3 to 4 months. In contrast, in the current study milk samples are collected from mothers feeding infants with age ranges from 20 days to 15 months. So there is comparatively more spread in the data due to extended ages groups [20]. For better separation, the number of principal components have been gradually increasd, an improved classification efficieny has been observed by increasing the number of PC's to 5 (considering first five PC's). For the purpose of visualization only the first 2 PC's has been considered because visualization in more than two dimension is difficult. The perfomace of the model has been evaluated using k-fold cross validation approach. For clarity purpose, loadings of the first four PC's are shown in figure (Fig. S1) as a supporting material.

### 3. Results and discussion

#### 3.1 Raman spectral analysis

Figure 1 shows vector normalized Raman spectra of overall milk samples from mothers of both types of genders. Similarly, Fig. 2 depicts the vector normalized mean Raman spectra of both types of milk samples. Throughout for demonstration, Raman spectra of milk samples of mother's feeding female infant are shown in red color whereas those recorded from milk of mother's feeding male infant are shown blue color. Raman spectra of both types of milk containing major and minor peaks at 838, 1065, 1120, 1150, 1250, 1290, 1440, 1515, 1640 and 1730  $\text{cm}^{-1}$ . Each Raman peak represents a specific molecular structure while its intensity shows the concentration of that specific molecule within the sample. Raman intensities of these peaks are almost comparable except at some shift position where slight variation can be seen as shown in Fig. 2. Figure 3 shows the vector normalized mean Raman spectra recorded from milk samples of both infant genders in four different age groups (same lactation stages were grouped).

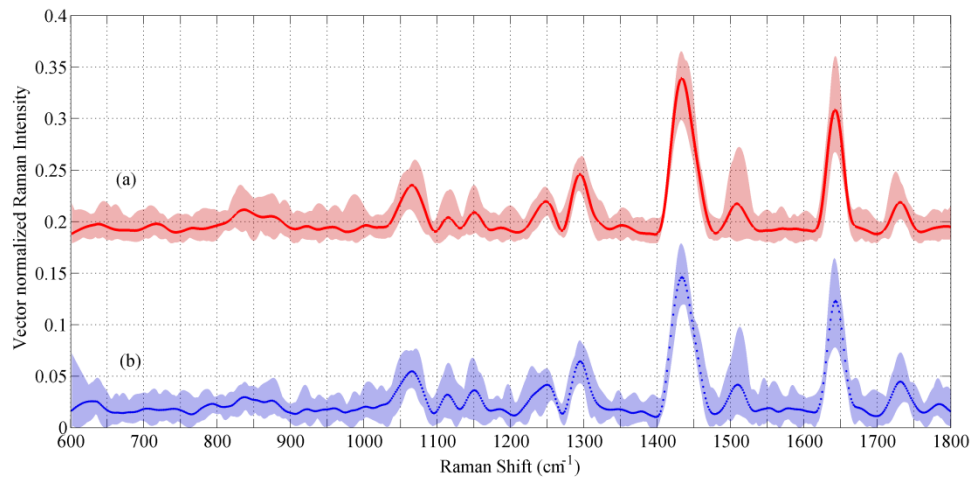


Fig. 1. Raman spectra of overall mothers milk samples of (a) female infant and (b) male infant genders. Shaded region depicts variations in the data while solid line depicts mean for both types of milk.

The characteristic Raman peak appeared at  $838\text{ cm}^{-1}$  and  $1120\text{ cm}^{-1}$  corresponds to saccharides [21]. Saccharides serve as prebiotics or antimicrobials influencing the bacterial communities in milk within the mammary gland. Although, these polysaccharides are not digested by the infants but feed the beneficial bacteria that live in the intestine which helps in fighting against infections [22, 23]. Raman peaks arises at  $1150\text{ cm}^{-1}$  and  $1515\text{ cm}^{-1}$  depicts the presence of carotenoids. Similarly, Raman peaks at  $1065\text{ cm}^{-1}$  and  $1440\text{ cm}^{-1}$  corresponds to triolein, oleic acid and palmitic acid which are strong fatty acids (cholesterol). Fatty acids are the main source of energy for the newborns during their first year of life [4, 24]. Fatty acids are responsible for the generation of docosahexaenoic acid (DHA) in newborns and are called smarter fats. Deficiency in DHA in the developing brain of infants leads to deficits in neurogenesis, neurotransmitter metabolism, and altered learning and visual function. Thus, sufficient evidence is available to conclude that maternal fatty acid nutrition serves as a source to supply DHA to the infant before and after birth, with short and long-term implications for neural function [9, 25].

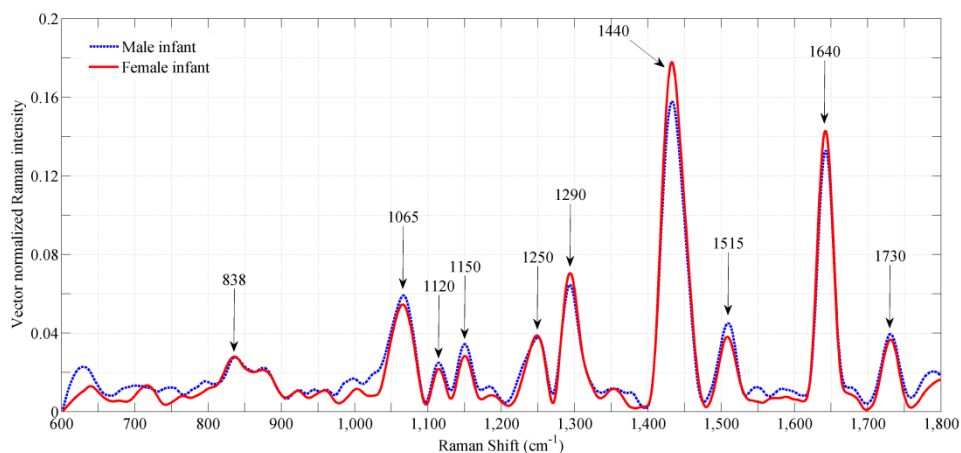


Fig. 2. Vector normalized mean Raman spectra of human milk samples having different infant genders.



Raman peak at  $1640\text{ cm}^{-1}$  corresponds to amino acids like tryptophan (Trp) which is an indispensable amino acid and is required for the biosynthesis of proteins, serotonin and niacin that helps in infant's normal growth [26–28]. Tryptophan and their metabolites regulate neurobehavioral effects such as appetite, sleeping-waking-rhythm and pain perception. Due to its high Tryptophan concentration, human breast milk protein provides optimal conditions for the availability of the neurotransmitter serotonin [10]. Raman peak at  $1250\text{ cm}^{-1}$  corresponds to cytosine, which is one of the important nucleotide (NT) for transferring genetic information, tissue development, immune defense development and sleep homeostasis in infants [29–32]. Similarly, Raman peak at  $1296\text{ cm}^{-1}$  corresponds to Phospholipids consisting of a variety of fatty acids having different head groups like Choline, Inositol and Serine. They are major part of the biological membrane and play an essential role in cell signaling, liver repair etc. Furthermore, Raman peak at  $1732\text{ cm}^{-1}$  arises due to the  $\text{C}=\text{O}$  stretching vibration of cortisone which normally passes from mother's blood stream in to milk.

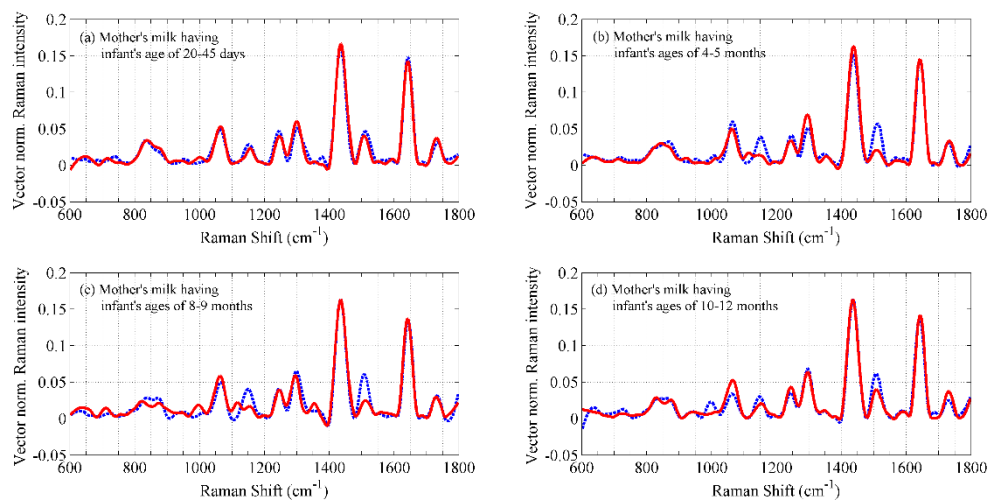


Fig. 3. Comparison of two types of milk samples showing different stages of lactation period. Dotted line represents milk feeding male infant and solid line female infants.

During lactation, milk composition and its volume modifies with time. Maternal diet, milk expression and lactation stage are the main factors involved in bringing variations within species and between species. Raman spectral results showed differences between both types of milk samples i.e. mother's feeding male and female infants based on spectral signatures. These results depicts that mother milk for female infants are richer in fatty acids, phospholipids and tryptophan than male infants; whereas, male infant received more carotenoids and sacchrides than female infants (Fig. 1 and Fig. 2). Additionally, at each stage of lactation, almost similar variations in concentration have been observed throughout (Fig. 3). Furthermore, the results shown in Fig. 3(a) also indicate that the contents of carotenoids in milk samples of both genders remains same in the earlier stage (first months).

### 3.2 Model evaluation

The performance of the model has been tested by generating confusion matrix using polynomial and Gaussian radial basis kernel. Confusion matrix is 2x2 matrix, presenting an event in the predicted as well as in the actual class by columns and rows respectively. As mentioned before PCA is unable to classify the given data well considering only two PC's, therefore first five PC's has been used for achieving better separation. The performance of SVM algorithm has been tested using k-fold (keeping  $k = 10$ ) cross validation method. Each time it divided an entire data set into  $k$  different sub-sets randomly i.e. 10 sub-sets. Every

time  $k-1$  sub-sets are used for training and raining one for testing purpose. An overall process is repeated  $k$ -times. In this way each sample has been predicated sequentially. Figure 5 shows hyperplane for 1 of 10 fold using polynomial kernel of order 2.

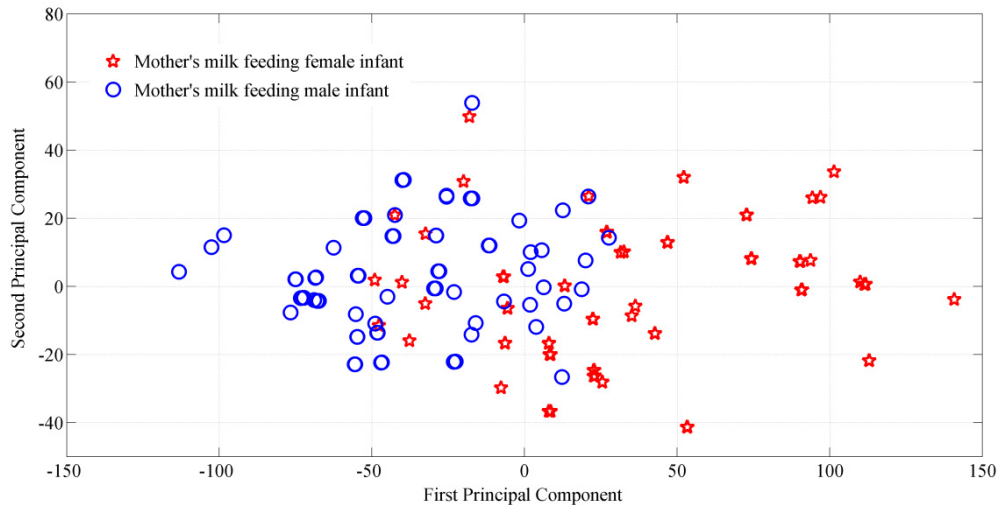


Fig. 4. Principal component scattered plot of human milk feeding male infant (open circle) and female infant (open star).

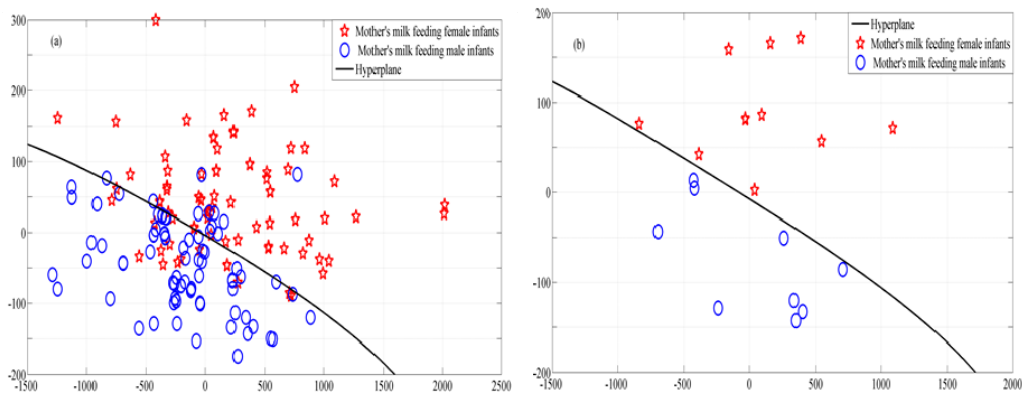


Fig. 5. polynomial kernel of order 2 separating boundaries for one of the 10-folds using the first two PCA transformed features; (a) training (b) classified.

For the purpose of presentation training and tested results are shown separately as Fig. 5(a) and Fig. 5(b) respectively. The overall model results using different kernel functions are shown in Table 1. From these results one can deduce that using polynomial as well as Gaussian radial basis kernel functions, SVM model effectively classify infant gender based milk samples. The performance of a model is generally evaluated in terms of accuracy, precision, sensitivity and specificity. Comparitively better performance has been achieved with polynomial kernel of order 2 with accuracy, precision, sensitivity and specificity of 86%, 88%, 85% and 88%, respectively.

Table 1. SVM- model results using polynomial and Gaussian radial basis kernel functions.

	Polynomial Kernel (Order-2)	RBF Kernel (rbf sigma-2)
True positive	84	81
False positive	11	14
True negative	80	82
False negative	15	13
Accuracy	86%	86%
Precision	88%	85%
Sensitivity	85%	86%
Specificity	88%	85%

#### 4. Conclusions

Raman spectroscopy combined with SVM differentiated milk samples of mother's having male and female infants. This differentiation is attributed to the variation in concentration of their fatty acids, phospholipids, tryptophan, saccharides and beta carotenoids. It has been found that milk sample of mother's having female infant is richer in fatty acids, phospholipids and tryptophan. In contrast, milk sample of mother's having male infant contains more carotenoids and saccharides. Additionally, at each stage of lactation, similar variations in concentration have been observed throughout. Principal component analysis and SVM further highlighted the differences between the two groups on the basis of differentiating features obtained from their Raman spectra. In the long run, these results may set new tradition and could be used as a reference for the preparation of infant gender based powdered formula milk which is recommended as a substitute of mother's feed. In this regard, further studies are in progress in our laboratory, considering the possibilities to explore gender specific differences within different species.

#### Acknowledgement

We are thankful to Muhammad Irfan and Mrs. Fatima Batool Scientific Assistants at Agri-Biophotonics Division, National Institute of Lasers and Optronics (NILOP), for assisting the present research work.

#### Disclosures

The authors declare that there are no conflicts of interest related to this article.